Big Data Analytics for Continuous Assessment of Astronaut Health Risk and Its Application to Human-in-the-Loop (HITL) Related Aerospace

Carolyn McGregor AM[1]
University Ontario Institute of Technology, Oshawa, Canada

Oleg Orlov[2], Roman Baevsky[3], Anna Chernikova[4], Vasily Rusanov[5],
Institute of Biomedical Problems Russian Academy of Sciences, Moscow, Russia

The man-instrumentation-equipment-vehicle-environment ecosystem is complex in aerospace missions. Health status of the individual has important implications on decision making and performance that should be factored into assessments for probability of success/risk of failure both in offline and real-time models. To date probabilistic models have not considered the dynamic nature of health status. Big Data analytics is enabling new forms of analytics to assess health status in real-time. There is great potential to integrate dynamic health status information with platforms assessing risk and the probability of success for dynamic individualized real-time probabilistic predictive risk assessment. In this research we present an approach utilizing Big Data analytics to enable continuous assessment of astronaut health risk and show its implications for integration with HITL related aerospace mission.

I. Introduction

THE human factor in various aerospace missions is a necessary component to consider as part of a complex man-instrumentation-equipment-vehicle-environment ecosystem. To assess the probability of success, various human in the loop (HITL) models have been proposed. One such approach is probabilistic predictive modeling (PPM), which, through repeatable experimentation is used to gain a better understanding of the role that various uncertainties play for humans working in the roles of planner's and operator's, as well as of the role of the human factor in various human-in-the-loop (HITL) related missions and situations[1]. However, to date, these approaches have not considered the health of the individual in these roles at the time of required activity and as such have not enabled personalized precision medicine approaches to the assessment of the probability of success of the proposed activity with a given individual at a given point in time. Health status of the individual has important implications on decision making and performance that should be factored into assessments for probability of success/risk of failure. In addition, new approaches to real-time health status monitoring utilizing Big Data analytics that consume physiological data from various sensors have great potential to provide dynamic analytics of human health state for integration with platforms assessing risk and the probability of success for dynamic individualized real-time probabilistic predictive risk assessment.

In space, monitoring human adaption in real-time as well as usual human health are important functions that are required to be assessed regularly. In recent years, new predictive approaches to health and working capacity assessment are developing in space medicine. It attaches prime importance not to the diagnosis but to assessment of

[1] Canada Research Chair in Health Informatics and Professor, Faculty of Business and IT, 2000 Simcoe St North, Oshawa, L1H7K4, Canada
[2] Director, Institute for Biomedical Problems, 76-A Khoroshevskoye shosse, 123007, Moscow, Russia.
[3] Professor and Principal Researcher, Laboratory of Autonomic Regulation of Cardio-Respiratory System, 76-A Khoroshevskoye shosse, 123007, Moscow, Russia.
[4] Senior Researcher, Laboratory of Autonomic Regulation of Cardio-Respiratory System, 76-A Khoroshevskoye shosse, 123007, Moscow, Russia.
[5] Senior Researcher, Laboratory of Autonomic Regulation of Cardio-Respiratory System, 76-A Khoroshevskoye shosse, 123007, Moscow, Russia.

health risks, including adverse functional conditions that may affect safety. The principles of prenosological (preclinical) diagnosis studies the human's functional states on the verge of normal and pathological conditions. Disease is considered as a result of violations of the body's ability to adapt to environmental conditions, as a failure of adaptation mechanisms. This is preceded by the stress of adaptation mechanisms that are aimed to mobilize the functional reserves. Then follow the steps of reduction and exhaustion of functional reserves, which are accompanied by premorbid conditions and insufficient adaptation. All these conditions, prior to the failure of adaptation mechanism and the emergence of disease, increase the so-called adaptive risk - the probability of the disease manifestations due to the reduction of adaptation capacities of the organism.

The simplest method to evaluate adaptation features is the HRV analysis. This method has been used in Russian space medicine since the first manned space flights to assess and predict the cosmonaut's health status. Changes of HRV have been reported in several cardiovascular and noncardiovascular diseases (myocardial infarction, diabetic neuropathy, myocardial dysfunction) and after specific interventions (drugs, exercise training, biofeedback techniques, workloads).

Probabilistic forecasting of health and working capacity is a fundamentally new approach to the safety of aviation and space flights. It essentially supplements and develops the traditional methods of assessing the quality and time of work operations. This novel approach is actively developed in the Russian space medicine and is based on a mathematical model of human functional states according to the heart rate variability (HRV) analysis in terrestrial studies and dozens of manned missions to the orbital stations "Mir" and the ISS[2]. Bernard Marr in his book "Big Data: Using SMART Big Data, Analytics and Metrics To Make Better"[3] mentioned the usefulness of HRV data from wearable smart devices in predicting of mortality, in understanding more about sleeping patterns and other physiological phenomena of adaptation.

In this research we present an approach utilizing Big Data analytics to enable continuous assessment of astronaut health risk and show its implications for integration with HITL related aerospace mission. This approach provides a systemic platform for the integration of several specific algorithms to measure and quantify health state.

II. Artemis

In prior work McGregor[4] created Artemis in partnership with researchers from the IBM TJ Watson Research Development Center through an IBM First-of-a-Kind award. Artemis, is a platform for online health analytics to enable real-time multi-patient, multi-diagnosis and multi-stream temporal analysis for real-time clinical management and retrospective research. The Data Acquisition acquires streams of data in real-time from the medical devices together with other periodical data from the Electronic Health Record. This data is then forwarded before it is stored for analysis in real-time within the Online Analysis component. Artemis utilizes IBM's InfoSphere Streams, for the real-time Online Analysis and then provides the ability to store the raw data together with the derived analytics within the Data Persistency component. Clinical decision support algorithms can run in real-time within the Online Analysis component as Infosphere Streams 'graphs'. Retrospective analysis of the stored data is performed by the Knowledge Extraction. The Artemis Architecture is presented in Figure 1 below.

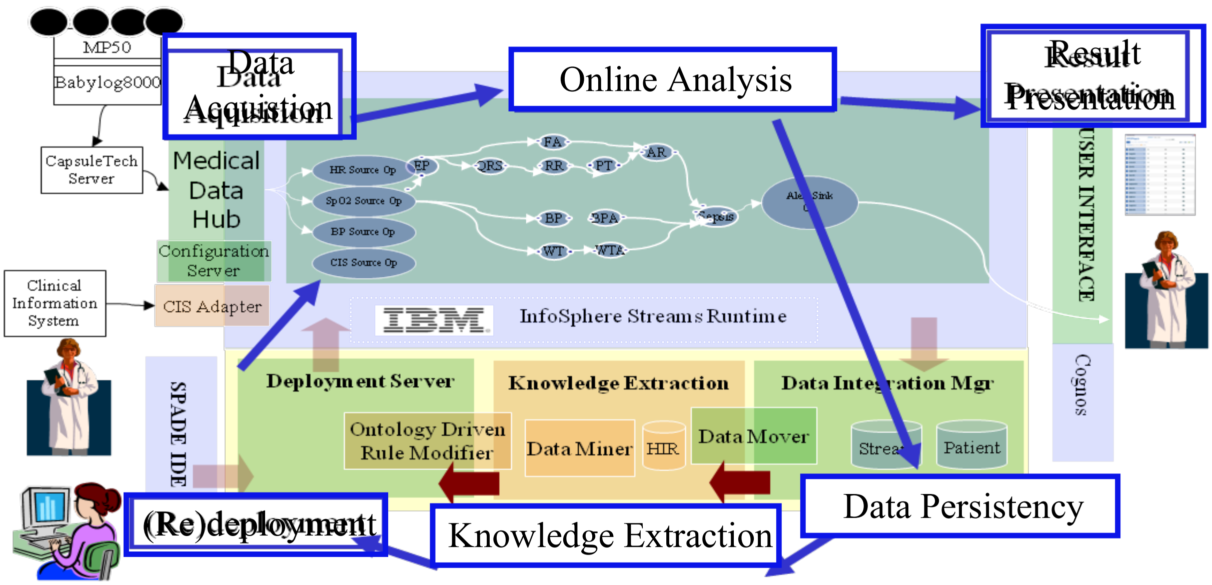

Figure 1. Artemis Framework[6]

Artemis processes and then stores the raw data and derived analytics data from multiple infants at the rate it is generated and has been used to support clinical research studies for neonatal conditions including: late-onset neonatal sepsis; apnoea of prematurity, in which the infant experiences pauses in breathing and reductions in heart rate and blood oxygen saturation; retinopathy of prematurity, which can result in permanent blindness; and pain[5]. Infosphere Streams graphs have been deployed for clinical decision support algorithms that can run in real-time for late onset neonatal sepsis, blood oxygen saturation analytics and apnoea of prematurity.

McGregor extended this work to propose a platform for real-time health analytics in space flight[7]. It includes components that would be onboard the spacecraft and components that would be on earth. In this platform, the data acquisition component receives physiological data from astronauts. Real-time analytics can be performed on-board the spacecraft within the Online Analytics component. Results from the analytics can be presented within the results presentation component and the raw data together with the results of the analytics would be stored in the Stream Persistency component. Retrospective discovery of knowledge from the data can be performed in the Knowledge Extraction component and new clinical decision support algorithms that result together with amended existing clinical decision support algorithms can be deployed and redeployed within the online analytics. Management of the deployment and redeployment process would be through established policies and procedures for clinical decision support systems for onboard the spacecraft. Data can be transmitted to the earth based Stream Persistency component where additional knowledge extraction research can occur simultaneously with the knowledge extraction research performed on the spacecraft. New clinical decision support algorithms generated on earch can similarly be deployed on the aircraft following the appropriate policies and procedures for deployment[7]. This is presented in Figure 2.

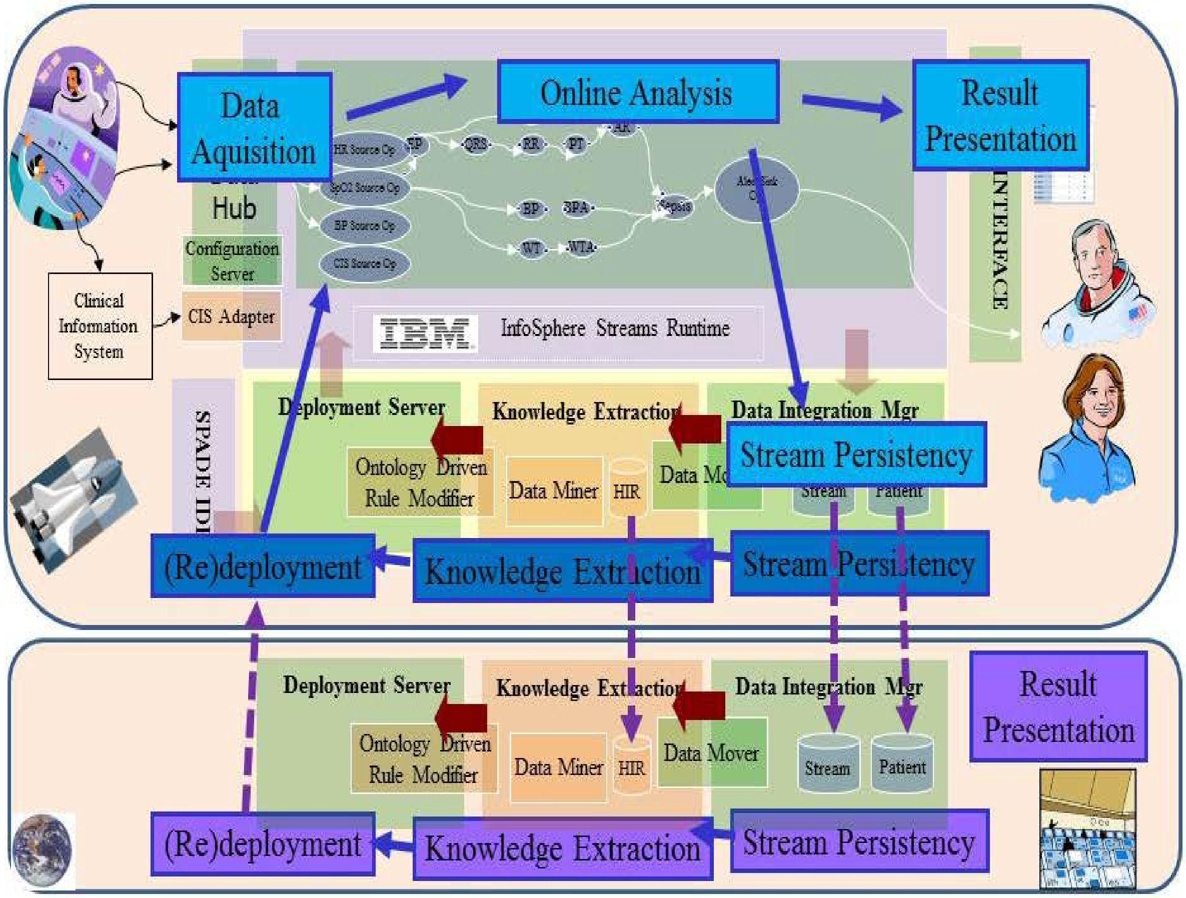

Figure 2. Online Health Analytics During Spaceflight[7]

III. Probabilistic algorithms for Functional State identification and Disease Risk Assessment

As noted in the Introduction section, a mathematical model of human functional states utilising heart rate variability (HRV) analysis has been the subject of prior Russian space medicine research[2].

The current Russian "Cosmocard" experiment that was started on the ISS as of September 2014 acquires electrocardiogram (ECG) data discontinuously from Russian cosmonauts on the International Space Station (ISS).

HRV data are obtained from ECG recordings (sample rate 1000 Hz). Analysis of HRV parameters are performed according to the standards of the European Society of Cardiologists and North-American Society of Electrostimulation and Electrophysiology. HRV characteristics in the time domain (SD, standard deviation; CV, coefficient of variation; RMSSD, root mean square of successive differences; pNN50, number of RR-interval pairs differing by more then 50 ms, SI, triangular index or stress index for characterization of the histogram) and in the frequency domain (TP, total spectral power of HRV; VLF, HRV power in the very low frequency range; LF, HRV power in the low frequency range; and HF, HRV power in the high frequency range) are calculated.

The canonical discriminant function is a linear combination of discriminant variables and satisfies certain conditions. It has the following mathematical representation:

$f = u + u X + u X + ... + u X$, (1)

where f – mean of canonical discriminant function for object m in group k;

X – mean of discriminant variable X for object m in group k;

u - coefficients that ensure execution of the required conditions.

Coefficients u for discriminant functions can be represented in standardized and raw form. For interpretation of the discriminant functions it is more interesting to consider standardized coefficients, which show the relative contribution of each variable.

Classification of each object and prediction of belonging to a particular class of objects is carried out by calculating a posteriori probabilities. In this case, the investigated object will be assigned to the class for which it has the greatest probability.

A classification system based on HRV analysis was developed[2] using data from reference group (192 volunteers women n=51, age range 38 to 62 years; men, n=141, age range 39 to 63 years). The most informative parameters for sufficiently high accuracy of recognizing specific functional states were mean heart rate (HR), the triangular histogram index (SI), pNN50 and the HF spectral power. The standard form of discriminant function equations for the first two canonic variables (L1 and L2) is shown below:

L1 = - 0,112*HR - 1.006*SI - 0.047*pNN50 - 0.086*HF;
L2 = 0,140*HR - 0.165*SI - 1.293*pNN50 + 0.623*HF.

Canonic variables were calculated from absolute values of HRV parameters. Analysis of standardized coefficients in these equations shows that SI has the highest weight in the first equation, while pNN50 and HF have the highest weight in the second equation.

Parameters L1 and L2 were referred to as coordinates of the phase plane configuring the space of functional states as shown in figure 1.

The functional states are situated on the phase plane in such a way that the physiological normal state is characterized by positive L1 and negative L2 values. The center is in the lower right quadrant of the phase plane. The remaining functional states are located in the other quadrants, i.e. prenosological (intermediate) state in the upper right, premorbid state – in the upper left and pathological state – in the lower left quadrant.

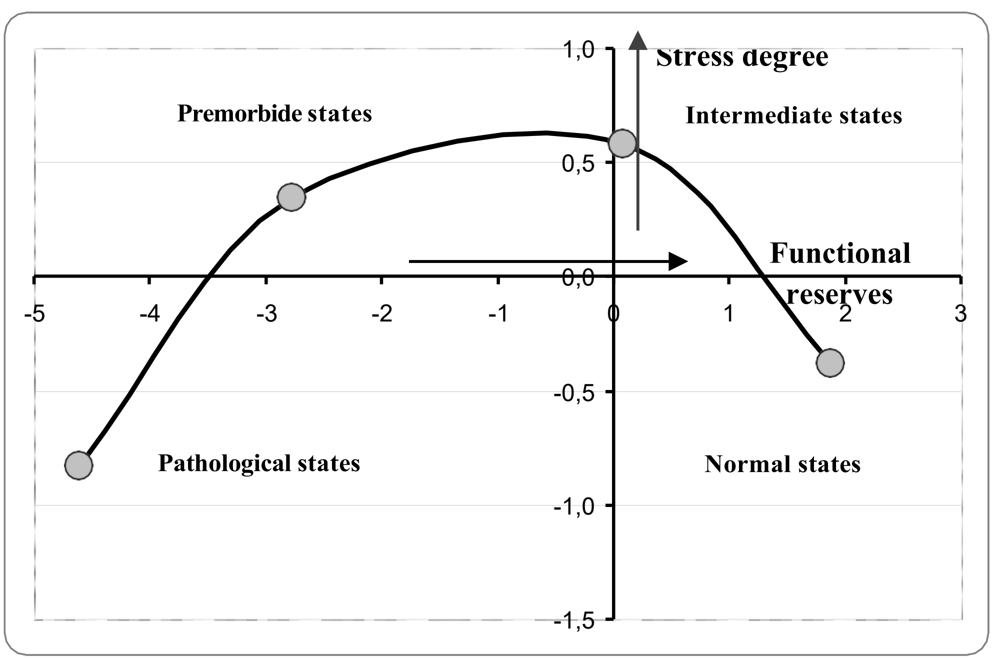

Figure 1. Geometric means calculated by using the canonical values L1 and L2 for each class of the reference group are shown.

One of the key limitations in this research to date is that the calculations of L1 and L2 are performed retrospectively and on Earth. In order to support real-time assessment of health state and to enable the probabilistic health predicting and working capacity assessment onboard the space craft new Big Data analytics approaches are required.

A second limitation of this research was that while the L1 and L2 were calculated based on data block intervals of 5 minutes, the number of L1 and L2 values in a 24 hour period, namely 288, was cumbersome to analyse manually and so this was down sampled to an average daily L1 and L2 score.

IV. Probabilistic Health Predicting in Real-Time

Currently IBMP (Russia) and UOIT (Canada) are collaborating in creating a new software system for the implementation of probabilistic health predicting and working capacity in members of space crews in real-time using the on-board computer. This will allow the crew of a piloted space object to independently monitor their functional conditions and together with the staff of the Mission Control Center make decisions about necessary preventive measures or correction of the flight program. When creating a new system we should additionally take into account many other data about work operations and environment. The first examinations of developing this new system have commenced during simulating terrestrial experiments with prolonged isolation in IBMP (Moscow).

All Probabilistic algorithms, developed by Institute of Biomedical Problems (IBMP), for identification of the functional state of the organism and for disease risk assessment, along with the Artemis platform for big data analysis, developed by McGregor, will be employed in this study.

This research extends that current Russian “Cosmocard" experiment. ECG data collected from this device from other related missions is used in this research.

In this research we utilize the principles of the platform proposed by McGregor as detailed in Figure 2. The data acquisition component received data from the Cosmocard device. Due to limitations of the Cosmocard device this data was acquired on the Cosmocard device and then downloaded and then replayed into the Data Acquisition component of Artemis to simulate real-time acquision of an ECG signal.

The probabilistic algorithms for functional state identification and disease risk assessment introduced in the prior section were reengineered to be calculated within an InfoSphere Streams graph that can be deployed within the Online Analytics component. The graph was designed to calculate all the metrics as introduced in the prior section and to also derive L1 an L2 and to perform this function of eack 5 minute block. The nature of the InfoSphere Streams environment enables tuples to be output each 5 minutes containing the set of metrics for that 5 minute block. This approach enables an easier data source for visualization of the movement of L1 and L2 over time in addition to the traditional down sampling to generate a daily average for L1 and L2.

For our initial testing of this approach we utilized data from an earth based research study called Luna 2015. In this research study six female crew simulated a moon mission. This data analysis was performed using the prior approach together with our new Big Data analytics approach and the results for each metric were compared to perform verification and validation of the generation of the metrics from the new Streams graph.

Results of the assessment of L1 and L2 and other related metrics from the Luna 2015 are outside the scope of this paper and will be reported in future publications.

Results presentation for this research we performed through the provision of the set of output tuples from the Streams graph to Microsoft Excel and the L1, L2 plot was constructed in Microsoft Excel.

V. Implications for Real-Time HITL Probabilistic Models

The ability to determine health state in real-time has great potential to be integrated within complex man-instrumentation-equipment-vehicle-environment models. This can be performed both offline for research and planning purposes to consider the impact of changing health state as a variable within those models. In addition, health state can be used for probabilistic modelling of a complex mission model for real-time current assessment of likelihood of success of the mission based on the current health state of the crew involved. Such systems are vital for the crew to have available especially when contact with Earth is not possible or is extensively delayed.

VI. Future Work

The use of the Cosmocard device places limitations on our ability to complete the analytics in real-time as the Cosmocard device is an acquire and store device and the data can only be extracted and analysed after the completion of data collection. While this is a limitation for our long term goal of providing real-time analytics, it did not prevent us achieving our goal of using the electrocardiogram data captured from that device and using it to simulate real-time collection and integration of that data within the new Big Data analytics environment for Online Health Analytics During Spaceflight.

In future work we aim to deploy this real-time analytics environment onto the International Space Station. We plan to develop additional clinical decision support algorithms. Using a different ECG acquisition device that enables real-time acquisition of ECG will support or long term goal of providing real-time health analytics.

VII. Conclusion

In this paper we have presented an approach utilizing Big Data analytics to enable continuous assessment of astronaut health risk and shown its implications for integration with HITL related aerospace mission. There is great potential for this approach to health risk assessment to be used offline in pre mission planning models together with in real-time for models to assess the potential of success in that moment based on the health status of the individuals involved. In our future work we will report on using our Big Data analytics approach for the analysis of data collected using the Cosmocard device during the Luna 2015 simulated moon mission. We will also replicate that work in future simulated missions prior to deployment of our environment on the International Space Station.

Acknowledgments

The authors wish to thank Anirudh Thommandram and Anastasiia Prysyazhnyuk from the University of Ontario Institute of Technology for their assistance with the translation of the probabilistic algorithms to an InfoSphere Streams graph for operation within the Big Data analytics platform based on Artemis. This research has been supported by the Canada Research Chairs program (#950-203427 and #950-225945), a National Science and Engineering Research Council (NSERC) Engage grant (Grant #461759) and the Canadian Foundation for Innovation (#203427).

References

[1]Suhir, E., “Human-in-the-loop: probabilistic predictive modelling, its role, attributes, challenges and applications,” *Theoretical Issues in Ergonomics Science*, Vol 16 No. 2, 2015, pp.99-123.

[2]Baevsky RM, Chernikova AG, Funtova II, Tank J,. “Assessment of individual adaptation to microgravity during long term space flight based on stepwise discriminant analysis of heart rate variability parameters,” *Acta Astronautica.*, 2011, Vol 69 No. 11-12, pp 1148-1152.

[3]Marr, B., “Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance,”, 2015, John Wiley & Sons.

[4]Blount, M., Ebling, M., Eklund, J. M., James, A., McGregor, C., Percival, N., Sow, D., “Real-Time Analysis for Intensive Care,” *IEEE Engineering in Medicine and Biology Magazine*, Vol 29, No. 2, 2010, pp110–118.

[5]McGregor, C., “Big Data in Neonatal Intensive Care,” *Computer*, Vol 46 No. 6, 2013, pp 54–59.

[6]McGregor, C., Catley, C., James, A., & Padbury, J. “Next Generation Neonatal Health Informatics with Artemis,” “*User Centred Networked Health Care,”*, IOS Press, 2011, pp. 115–119.

[7]McGregor, C., “A Platform for Real-time Online Health Analytics during Spaceflight,” *IEEE Aerospace Conference*, Big Sky, MT. 2013